DOCUMENT RESUME

ED 128 416 TM 005 600

AUTHOR Rodriguez-Brown, Flora V.; And Others
TITLE Longitudinal Design Considerations for the Evaluation of Bilingual Programs.
PUB DATE [Apr 76]
NOTE 18p.; Paper presented at the Annual Meeting of the American Educational Research Association (60th, San Francisco, California, April 19-23, 1976)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Academic Achievement; *Bilingual Education; Bilingual Students; Community Characteristics; Comparative Testing; Data Collection; Evaluation Methods; Language Programs; *Longitudinal Studies; Norms; Parents; Program Content; *Program Evaluation; *Research Design; School Districts; Statistical Analysis; Student Characteristics; Student Evaluation; Test Validity

ABSTRACT

Five elements necessary for the rigorous evaluation of bilingual programs are discussed: (1) careful collection of meaningful baseline data from selected subjects; (2) the identification and development of instruments to measure key variables, such as context and student characteristics; (3) the identification of treatment characteristics, such as curriculum design and materials, or the method of language usage; (4) the establishment of longitudinality; and (5) the interpretation of results in implementable terms that are meaningful to teachers, policy makers, and researchers. Other considerations discussed are the validity of available tests, data collection and management, and statistical analysis of the data. (BW)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED128416

Longitudinal Design Considerations for the Evaluation of Bilingual Programs

Flora V. Rodríguez-Brown
Andrew D. Cohen
Center for Applied Linguistics
&
Illinois Bilingual Evaluation Center

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Taweewat Pitayanon
Illinois Bilingual Evaluation Center

Thomas C. Ripley
IIT Research Institute
Chicago, Illinois

Paper presented at the 60th Annual Meeting of the American Educational Research Association, in San Francisco, California, April 19-23, 1976.

TM005 600

Introduction

Bilingual education is enjoying its first decade of prominence in the United States. In 1963, Dade County in Florida started a public school Spanish-English bilingual program for Cuban Americans and Anglos. In 1967, the Bilingual Education Act was added to the Elementary and Secondary Education Act of 1965. Federally-funded Title VII bilingual education programs began in 1968. More recently, states have passed legislation to fund bilingual programs (Swanson, 1974; U.S. Commission on Civil Rights, 1975).

At a time when the implementation of bilingual programs has reached such a peak, the evaluation of programs has lagged far behind. Despite millions spent on the development of programs, the United States experience to date has yielded few meaningful insights into various aspects of program design (Troike, 1974; U.S. Commission on Civil Rights, 1975; Ramirez et al., 1975). Reasons for this lack of hard data include the following:

(1) It is hard, if not impossible, to obtain meaningful research results from pilot programs that are constantly undergoing modification, presumably for the better. Even if summative results are obtained, the researcher is hard put to give a label to a particular treatment, since it is in such a state of flux.

(2) There has been such a pressing need for formative evaluation of project-oriented goals, specifically behavioral objectives contained in the curriculum, that no time has remained for evaluating other things.

(3) Until recently there have not been adequate assessment instruments particularly bilingual ones, and even now much test development and norming are called for.

(4) Political threats to bilingual schooling have almost forced evaluation reports to be public relations documents.

(5) Evaluators have tended to be persons unfamiliar with particular needs and characteristics of bilingual education.

The "fledgling progam" reason should no longer apply, since bilingual projects nationwide now have more stability, as a result of a gradually growing accumulation of experience, methods, and material. But if bilingual education is to continue to advance, better and more meaningful evaluation is necessary.

With respect to project-centered instructional objectives, more than ever before there is a need to entertain the larger questions as well. Tucker and d'Anglejan (1971) question whether "self centered" project goals such as meeting specific teaching objectives, are valid criteria for evaluating the success or failure of a program (e.g., 75% of the children can

answer 90% of the questions in a certain section of a book). Whether or not such criteria are valid, there is more to formative evaluation, such as investigation of the following areas (adapted from Saville and Troike, 1971):

(1) The teaching techniques that prove most successful in different situations (grouping, sequencing and pacing of materials, and correction procedures).

(2) The effect of program design (e.g., partial or full bilingual schooling using a concurrent, dual language, or alternate days approach to instruction).

(3) The effect of teacher training and patterns of staff utilization.

The lack of adequate instruments is still a problem, though not as severe as 7 years ago when federally-funded program evaluations were initiated. Yet evaluation must proceed even if the most appropriate instrument is not available. Recently even some widely used standardized instruments, such as the Peabody Picture Vocabulary Test and the Cooperative Primary Test of Reading, have been subject to criticism (Cicourel et al., 1974). We seem to be entering an era in which ethnomethodological scrutiny of tests and of individual test items will be common practice.

Finally, it would appear that bilingual schooling is here to stay, at least for the foreseeable future. Thus, evaluation reports should reflect more than a morass of tabular data and a scattering of carefully selected and tentatively-or even ambiguously-worded findings. Instead, the findings should reflect strengths as well as weaknesses, and even more important, should be designed so as to provide feedback to aid in the ongoing improvement of program practices. It is regretable that the tendency to avoid measures which might produce negative results has all but precluded the possibility of learning from project deficiencies (Berman and McLaughlin, 1974).

Given the current developments in the field of research, there appear to be few obstacles to conducting sound, rigorous evaluation of bilingual programs. Such evaluation would reflect the following elements:

(1) Careful collection of meaningful baseline data from selected subjects.

(2) The identification and development of instruments to measure key variables.

(3) The identification of treatment characteristics.

(4) The establishment of longitudinality.

(5) The interpretation of results in implementable terms that are meaningful to teachers, policy makers, and researchers.

Whereas the research literautre on bilingual schooling is generally lacking rigorious longitudinal evaluations, several such investigations have been conducted in Redwood City (Cohen, 1975) in Culver City (Cohen, 1974), and in Montreal (Bruck et at., 1975).

Rationale

The purpose of this paper is to present some considerations when developing a longitudinal design for the evaluation of bilingual programs.

According to Goulet (1975), the longitudinal design method requires the testing of samples with the same birth date or alternate samples who are at the same grade level at different times. This design is useful for between-subjects and within-subject (i.e., repeated measurements) testing procedures. The requirement of longitudinality is met when the subjects, at the same grade level, are tested at two or more points in time. This design could be called a within-subject one.

We offer six specific motivations for employing a longitudinal design. These are:

1. Students are usually enrolled in a bilingual program for more than one year.

2. The effects of these programs become evident over a longer period of time than one year.

3. A longitudinal design may permit the evaluation of program development.

4. The bilingual students individually and as cohorts can be used as a comparison group.

5. The data necessary for a longitudinal analysis are usually collected by repeated one-shot designs and only some careful forethought and data management are required to implement a longitudinal model.

6. The wealth of information available in the data collected according to a longitudinal design, answer many of the reoccuring questions relating to progress in pedagogy.

If we were to use a cross-sectional design, where samples of different ages are tested at the same point in time, we may be required to assume that the effects of schooling for children in comparable grades are the same irrespective of the year in which the children are enrolled. In such a changing world as the one we live in today, there exists a lot of danger in making this assumption.

With respect to the one-shot design, often used by school districts to report their gains, success, etc., in a year, we feel that this design does not give enough information about the process of bilingual schooling. A design of this type will tell us about the product of one year of schooling, but very little about the process.

It is expected that a longitudinal evaluation design will give us more detailed information about the process and the effect and in a long term the product of bilingual schooling. A longitudinal study may show some developmental trends followed by children under the bilingual education treatment which differ or compare with those of children attending regular classrooms. By following the same subjects through the years, it is possible to control more for treatment and general school experiences which may affect children.

In the case of longitudinal design, we will get information as to student gains and achievements each year. At the same time, we are able to find out more about the process of bilingual schooling by comparing the data for the years the child attended the program. We can use the student as his own control and find his achievement, cognitive, etc., gain each year and we can make comparisons within the group and even at the cross-sectional level, if necessary.

The implementation of a longitudinal design requires a lot of planning, since it is a long-range process. First considerations should be taken about treatment and control groups. The problem of attrition should be taken into account when deciding the size of the two groups. Economical considerations are important, handling of the data and maintaining a data base become an extra expenditure. A longitudinal design will require the collection and storing of much ethnographic data then required in a one year study. It is important to control and know details about extra factors which could influence schooling in one way or another.

Before considering the possible designs which can be used, a word regarding our philosophy of evaluation is in order. The definition of evaluation is disputable and has been attended by a multitude of experts over the years. The definition employed in this paper is that program evaluation is the assessemnt of the program's worth. This, of course, implies a relative source of continuous debate. However, one must acknowledge some standard, whether it be a comparison group, national norms, or individual student histories. If a program does not incorporate such a standard in its objectives at least implicitly, the program's objectives must be judged inadequate. Thus the design presented does include a comparison group.

The Design

The advantage of a longitudinal model, in addition to permitting the assessment of program outcomes that may occur in the future, is that the program's impact may be seen as a discontinuity in the students achievement history. Although marked progress in student achievement relative to his history is not proof of the program's impact, occurrence of such progress in a large proportion of the program's students would be welcomed.

The particular evaluation model (analysis) that should be adopted is highly dependent on the existence and equivalence of a comparison group. Horst (1975) has proposed five models based on comparison group equivalence. These are:

- posttest comparison with matched groups
- covariance analysis

- special regression analysis
- general regression analysis
- norm-referenced

Horst (1975) persents these models and their strengths and weaknesses and refers to much of Campbell's work (1974, 1970, 1966). However, some comments regarding longitudinal and bilingual aspects are in order.

The norm-referenced model is not readily applicable to bilingual programs, because many of the measures used in these programs have not been normed. Also, objectives of bilingual programs. Thus, the norm-referenced model is not frequently applicable. This necessitates the use of a comparison group.

Goulet (1975) presents a nice figure to illustrate the difference between cross sectional, longitudinal and time lag models. As seen in Figure 1, C represents a cohort which could be students with the same birth date, or students who started the program in the same year. A represents the age of the group or the number of years in the program. T, of course, represents different times. Note that if measures are made for every cell in this matrix, a number of reflexive comparison groups can be isolated. For instance, in the longitudinal vector each group can be compared against itself at different times. Along the time-lag vector, each cohort can be compared against another who has had the same treatment exposure but over a different time period. Along the cross sectional vector, cohorts can be compared against each other for different treatment exposure but at the same time.

It is quite obvious that each of these comparisons are confounded in different ways. However, the cohort effects may be insignificant or partialed out to permit estimates of the trend effects of bilingual program from the time-lag comparisons. These trends will provide answers to the question "is the bilingual program improving over time or not?" However, the question, "is this program better than others?", still remains. This question again brings us back to the comparison problem.

Model 1, the posttest comparison with matched groups, is preferred from the evaluative standpoint. The children are paired in terms of pretest measures and random assignment of one member of each pair is assigned to the treatment group. The most evident practical drawback of this model is the random assignments process. Many administrators of special programs contend that the "neediest" students must receive the special instruction. This fact better than the existing ones, why aren't all students receiving this program? If it is a matter of money only, funds used for an evaluation of a proven program should be placed into the program operation coffers so that at least a few more students can enjoy the benefits of the special program.

More often than not, bilingual programs are experiments that may or may not enhance student achievement. A philosophy that all educational programs offered are the best they can be given various contextual constraints and pedagogical knowledge is requried. New programs including bilingual are experiments and random assignment of students to programs in no way violates

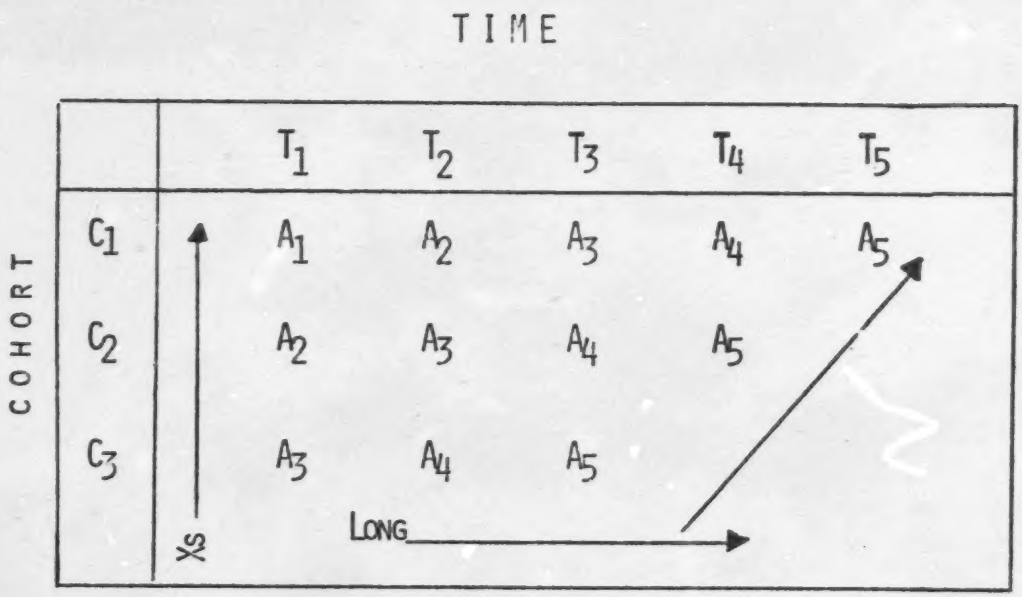

FIGURE 1. REPRESENTATION OF CROSS-SECTIONAL, LONGITUDINAL AND TIME LAG DESIGNS

the rights of the neediest child to his equal educational opportunity. With these arguments, randomization is justifiable and where possible matched groups are urged.

Where matching is unfeasible or impossible, the analysis of covariance model is appropriate. Again, random assignment is assumed and parallelism of regression lines between groups should be tested. Where departures from parallelism exist, analysis of covariance tends to underadjust the comparison group posttest results.

Where randomization is ruled out as a method for assigning students to groups, two special regression models, the Regression Projection Model (Tallmadge and Horst 1974) and the Regression Discontinuity Model (Campbell and Stanley 1963) are offered. These are aptly presented in (Horst 1975) and are only recommended when restrictions are placed on randomization.

Too many times evaluation of programs are requested after they have been implemented. In these instances there may be little hope that methodological development of a comparison group was undertaken. In these situations the general regression model is the only feasible approach. This model is the most flexible but tends to underadjust, particularly when pretest and posttest correlation is low. However, this type of error results in conservative inferences concerning program effects. This model is often used in retrospective studies when no methodological assignment was employed.

Superimposing a time series framework on these models permits comparisons of the time effects of programs. Programs can be compared regarding their longitudinal or trend effects, their robustness to varying cohorts over time, and their relative maturation.

The recommended longitudinal design is not longitudinal in the sense of Goulet (1975), but is rather the composite design of Figure 1 superimposed on Model 1, the posttest comparison model for matched groups. Superimposing Figure 1 on the other models is decreasingly less desirable. (The use of Model 4, the general regression model, is defined as a retrospective design). The above approach may seem somewhat grandiose, requiring a good deal of data over a period of years. This however, is not the case. These data are usually collected each year and the application of this approach just requires some thought and careful management of data collected over time. Such management is the crux of successful evaluation.

Data Dilemmas

Conducting longitudinal bilingual evaluations require one to contend with a number of dilemmas. Some of these dilemmas are unique to bilingual programs while others are common to educational programs and change measures in general. The dilemmas that will be considered are:

- variable selection
- validity
- collection

- management
- attrition and missing data
- comparability
- analysis

The variables selected for measurement are determined by two primary considerations. These are:

- What are the goals or objectives of the school district or educational organization?
- What controls are needed to provide meaningful comparisons among programs?

From answers to the latter question, the set of independent variables can be developed. With this idea in mind we have defined a series of independent variables that we feel are relevant to a longitudinal study of bilingual schooling. These variables are loosely defined and only attempt to touch on factors that should be considered for an evaluation design. The importance of each factor as a source of variation that must be controlled, must be made in light of the specific evaluation settings.

We have divided these variables into three groups or categories.

I. Contextual Variables

II. Student Variables

III. Treatment Variables

A list of these variables under each category is given below.

I. Contextual Variables

A. School district characteristics

1. size
2. resources
3. ethnic composition
4. degree of integration
5. SES composition

B. Community characteristics

1. density
2. SES composition
3. degree of integration
4. occupational make-up
5. political involvement including involvement in school district
6. educational attitudes

C. Parent characteristics

1. schooling
2. occupation
3. ethnicity
4. attitudes
5. involvement
6. SES status
7. dominant language
8. children-home life

II. Student Variables

A. Physical characteristics

1. sex
2. size
3. health (physical handicaps)
4. age

B. Education

1. level of schooling
2. years of schooling
3. schooling characteristics

   a. grades (marks)
   b. continuity
   c. special program (other bilingual programs)

4. attitude toward school and education
5. dominant language

   a. reading
   b. speaking
   c. listening

6. achievement

   a. home language context
   b. English context

C. Peer relations

D. Language association

1. years in U.S.
2. age of first association with English
3. duration and time history with association
4. intensity of association

III. Treatment Variables

A. Setting

1. school characteristics
2. classroom characteristics
3. other programs employed

B. Program characteristics

1. size
2. staffing characteristics
3. personnel relations
4. selection criteria
5. curriculum

    a. design
    b. organization
    c. role of culture

6. materials
7. language usage

    a. allocation to subjects
    b. amount
    c. method
    d. peer usage
    e. student teacher usage

As one can see the list of independent variables is lengthy to begin with and this list is by no means complete. Because of the large number of potential, independent variables, the advantage of random assignment of students to groups becomes apparent. When Model 1 is employed only treatment variables need to be considered. Of course pretest measurement must also be used as a matching criterion.

The answers to the question concerning the goals and objectives of the program should define the dependent variables to be measures. In our consideration of dependent variables, we will try to cover general areas of concern. As we mentioned at the beginning of this paper, there are not many good tests available for the measurement of common bilingual objectives. We feel that any test chosen should measure the objectives of the program and should not be used only because of its availability. This may mean the development of tests specifically for a given bilingual program.

Some usual measures that reflect bilingual objectives are:

I. Achievement

A. Reading

1. home language
2. English

B. Academic and cognitive ability

1. in home language
2. in English

C. Math

D. Science

E. Social Studies

F. Language ability

1. listenting
2. reading
3. writing
4. speaking

II. Language Dominance and Parity

III. Affective Development

A. Self-esteem

B. Self-concept

C. Attitudes

One of the most prominent dilemmas in evaluation is the validity of measures on the independent and dependent variables. Some considerations in this area are:

1. The measures should at least be reliable.

2. When normed scores are used the norming group should have similar characteristics to the children being measure (i.e., language, culture, social economic status, etc.), and measurement should be made on the treatment groups at the same time during the school year that the norm group was tested.

3. If the instruments have parallel forms in English and a second language, the forms should have been adapted and not just translated.

4. The language use to obtain measures should only include the language the children use at their development level.

5. The measures should be culturally sensitive.

6. Administration and scoring should be straightforward and objective.

Some unique problems of validity occur when pretest and posttest differences or more complex measures of change are used. These problems are aptly described by Bereiter, Webster, and Lord (1963). Namely, these problems include the regression effect paradox, the reliability of estimated change, the effect of change on group heterogeneity, spurious correlation between change and some other variable. Most of these problems arise from measurement error on dependent variables. By computing change, these measurement errors confound the effects of treatment and contextual variables. Model 1 or 2 recommended for use in the longitudinal design do not encounter these models cannot be employed, care should be made in interpreting evaluation results and references cited in this paper should be reviewed. Encouraging results cited by Richards (1975) demonstrated through simulation that, all estimates involving pretest posttest differences measure school impact with reasonable accuracy. It is improtant to measure

change over the entire course of learning, however, and not just over the later stages of learning. The correlations between change scores and other school characteristics reflect with reasonable accuracy, the relationships between those characteristics and impact but consequently will be large only when the underlying relationships are substantial. Simple gain scores measure the true situation about as accurately as other change estimates, are easier to compute, and probably are more meaningful to non-researchers.

In this study, students were assigned to schools both randomly and non-randomly.

Another assault on validity of longitudinal studies occurs by the mere fact that the time period over which measures are made is greater than in one-shot design. Thus, changes can occur in the time-dependent contextual variables. For this reason, these variables should be measured on more than one occasion along with the dependent measures. Ideally, such measures should be made concurrently.

One of the most straightforward tasks of longitudinal evaluation is data collection and management. Yet, this task is usually the one that requires the most effort and is usually poorly done, resulting in invalid evaluation. Competency of data gatherers and their managers is mandatory. Some considerations in collecting and managing the data are:

1. Data should be maintained on a per student basis.

2. All students should be given one and only one unique identification number and this should be recorded on all information collected.

3. A computerized data base should be developed where possible to organize and maintain the data.

4. Sorting of students by informative identification numbers can provide an easy to use directory.

5. Meaningful identification numbers can be produced by using indicators of student characteristics such as the school, program, and section he is enrolled in, his birth year, year he entered the program, and grade he entered the program, etc.

6. Computer routines for validity checking should be incorporated into the data management system.

7. Simple edit, sorting, and merging routines should be set up for production use.

8. All data collection and management activities should be the responsibility of one person. This will avoid confusion and misinformation that normally occur when many data gathering activities are undertaken.

9. Many of the data management duties require the technical expertise of a good computer services staff which has some knowledge of statistical software that may be applied for evaluation.

Longitudinal designs are more susceptible to missing data problems through attrition and other reasons. All efforts should be made to avoid missing data. Where such problems do occur, there is very little elegant recourse. Some possible compensating steps which are not without bias are:

- Exclude records that have missing data.
- Estimate missing data from regression equations developed from available data. (In this case, the 95% confidence intervals could be use rather than the point estimates and appropriate maximum likelihood regression techniques could be applied to handle the mixed data forms, that is point and interval values.
- Scale down the evaluation to include only that set of variables for which complete data are available.

As stated each of these approaches are biased. The degree to which they can be applied depend on the data at hand.

Another primary dilemma of longitudinal evaluation and specifically bilingual evaluation is the comparability of measurement instruments over time. Tests in bilingual education that are related over various educational levels are scarce. These tests are usually not normed and thus one level is not related to others. The concept of grade equivalents has not been applied to bilingual measures. Thus measures of student progress over time may have to be developed before meaningful trend analysis can be performed. This is a major problem since a great deal of time, effort, and expertise must be employed to develop tests that measure the same concepts at various levels. The authors have no good suggestions for handling this problem other than to start from scratch. The selection of location and scale must be done carefully so that treatment effects are not masked over time. Thus local standarization should be done on the pooled measures across which comparisons are to be made.

The last consideration is that of analysis. Since the longitudinal model offered here is that of repeated measures, univariate analysis may not be appropriate and multivariate analysis may be necessary (Book 1963). Thus, the covariance matrix of student x time must be examined before appropriate analysis is selected. Methods for selecting the appropriate analysis and its subsequent application are aptly explained by Bock (1963, 1968, 1975). Computer software such as Jeremy Finn's Multivariance package should be sufficient to handle such designs.

Conclusion

This paper has attempted to touch upon many of the considerations one must take when planning a longitudinal bilingual evaluation. In doing so, it has attempted to cite some useful references which may explicate possible dilemmas. The authors expect to elaborate on the considerations in book or monograph form in the near future.

Bibliography

Balasubramonian, K., Seelye, H.N., & De Weffer, R.E., Do bilingual education programs improve English language achievement?: A report on an Illinois experiment. Mimeo, Illinois Bilingual Education Service Center, Mount Prospect, Illinois, 1974.

Berman, P., & McLaughlin, M.W., Federal programs supporting educational change, vol. 1: A model of educational change. Prepared for the U.S. Office of Education, H.E.W., R-1589/1-HEW, Santa Monica, California: Rand Corporation, 1974.

Bock, R.D., "Multivariate analysis of variance of repeated measurements," in C.W. Harris (ed.), Problems in Measuring Change. Madison: University of Wisconsin Press, 1963.

Bock, R.D., & Peterson, A., "Matrix operations subroutines for statistical computation." University of Chicago: Department of Education Statistical Laboratory, Research Memorandum No. 7, 1967.

Bock, R.D., Multivariate Statistical Methods in Behavioral Research. McGraw-Hill, 1975.

Bruck, M., Lambert, W.E., & Tucker, G.R., Cognitive consequence of bilingual schooling: The St. Lambert project through grade 8. Paper presented at the 9th Annual TESOL Convention, Los Angeles, March 4-9, 1975.

Campbell, D.T., Making the case for randomized assignment to treatments by considering the alternatives: Six ways in which quasi-experimental evaluations in compensatory education tend to underestimate effects. Draft report # Pre627, Northwestern University, Evanston, Illinois, 1974.

Campbell, D.T., & Erlebacher, A.E., How regression artifacts in quasi-experimental evaluations can mistakenly make compensatory education look harmful. In J. Hellmuth (ed.), Disadvantaged Child. Vol 3. Compensatory Education: A National Debate. New York: Brunner/Mazel, 1970.

Campbell, D.T., & Stanley, J.C., Experimental and quasi-experimental design for research and teaching. In N.L. Gage (ed.), Handbook of Research on Teaching. Chicago: Rand McNally, 1963.

Chicago Board of Education. Final evaluation report of State-supported bilingual education program for fiscal 1974. Department of Government Funded Programs, Chicago Board of Education, City of Chicago, 1974.

Cicourel, A. et al., Language use and school performance. N.Y. Academic Press, 1974.

Cohen, A.D. The Culver City Spanish immersion program: The first two years. The Modern Language Journal, 1974.

Cohen, A.D. A socialinguistic approach to bilingual education: Experiments in the American Southwest. Rowley, Mass.: Newbury Hourse. 1975.

Cohen A.D., & Laosa, L.M., Approaches to second language instruction: A reserch model. Paper presented at the American Educational Research Association Meeting, Washington, D.C., March 31 - April 3, 1975.

Finn, J.D. MULTIVARIANCE: a fortran iv program. National Educational Rescources, Inc., Ann Arbor, Michigan, 1972.

Goulet, L.R., Longitudinal and time-lag designs in educational research: an alternate sampling model. Review of Educational Research. Fall 1975 vol. 45 no. 4 pp. 505-523.

Horst, D.P., Tallmadge, G.K., & Wood, C.T., A practical guide to measuring project impact on aruswnr achievement. Under contract of HEW, RMC Research Corporation 1975.

Lord, F.M., Elementary Models for Change in C.W. Harris (ed.), Problems in measuring change. Madison: University of Wisconsin Press, 1963.

Phillips, J.M. Code-switching in bilingual classrooms. M.A. Thesis, California State University, Northridge, 1975.

Ramirez III, M. et al. Spanish-English bilingual education in the United States: Current issues, resources and recommended funding priorities for research. National Institute of Education Contract =NIE-C-74-0151. East Santa Cruz, California: Systems and Evaluations in Education, 1975.

Saville, M.R. & Troike, R.C. A handbook of bilingual education. Revised edition, Washington, D.C.: Teachers of English to Speakers of Other Language, 1971.

Seelye, H.N. & Balasubramonian, K. Accountability in educational reform programs through instrumentation analysis and design variation: Evlauations of cognitive growth in Illinois bilingual programs, 1972-73. Mimeo, Illinois Bilingual Education Service Center, Mount Prospect, Illinois, 1973.

Sinclair, J.M. &Coulthard, R.M. Towards and analysis of discourse: The English used by teachers and pupils. London: Oxford University Press, 1975.

Tallmadge, G.K. The development of project information packages for effective approaches in compensatory education. Los Altos, California: RMC Research Corporation, December, 1974 (Technical Report No. UR-254).

Tallmadge, G.K. & Horst, D.P, A procedural guide for validating achievement gains in educational projects (Revised). Los Altos, Calif.: RMC Research Corporation, December, 1974 (Technical Report No. UR-240).

Troike, R.C. International perspective in bilingual bicultural education. Annual International conference on Bilingual Bicultural Education, New York City, N.Y., May 15-18, 1974.

Tucker, G.R. & d'Anglejan, A. Some thoughts concerning bilingual education programs. The Modern Language Journal, 1971, 55.8, 491-493.

U.S. Commission on Civil Rights. A better chance to learn: Bilingual bicultural education. U.S.C.C.R. Clearinghouse Publication 51, May, 1975.